Reinforced Self-Training (ReST) for Language Modeling

Year: 2023 | Citations: 367 | Authors: Caglar Gulcehre, T. Paine, S. Srinivasan, Ksenia Konyushkova, L. Weerts

---

Abstract

Reinforcement learning from human feedback (RLHF) can improve the quality of large language model's (LLM) outputs by aligning them with human preferences. We propose a simple algorithm for aligning LLMs with human preferences inspired by growing batch reinforcement learning (RL), which we call Reinforced Self-Training (ReST). Given an initial LLM policy, ReST produces a dataset by generating samples from the policy, which are then used to improve the LLM policy using offline RL algorithms. ReST is more efficient than typical online RLHF methods because the training dataset is produced offline, which allows data reuse. While ReST is a general approach applicable to all generative learning settings, we focus on its application to machine translation. Our results show that ReST can substantially improve translation quality, as measured by automated metrics and human evaluation on machine translation benchmarks in a compute and sample-efficient manner.